

## Durham Research Online

---

### Deposited in DRO:

30 October 2020

### Version of attached file:

Accepted Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Hu, Shanfeng and Shum, Hubert P.H. and Liang, Xiaohui and Li, Frederick W.B. and Aslam, Nauman (2021) 'Facial reshaping operator for controllable face beautification.', *Expert systems with applications.*, 167 . p. 114067.

### Further information on publisher's website:

<https://doi.org/10.1016/j.eswa.2020.114067>

### Publisher's copyright statement:

© 2020 This manuscript version is made available under the CC-BY-NC-ND 4.0 license  
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

### Additional information:

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# Facial Reshaping Operator for Controllable Face Beautification

Shanfeng Hu<sup>a</sup>, Hubert P. H. Shum<sup>b,\*</sup>, Xiaohui Liang<sup>c</sup>, Frederick W. B. Li<sup>b</sup>,  
Nauman Aslam<sup>a</sup>

<sup>a</sup>*Northumbria University, Newcastle upon Tyne, UK, NE1 8ST*

<sup>b</sup>*Durham University, South Road, Durham, UK, DH1 3LE*

<sup>c</sup>*Beihang University, Beijing, China, 100191*

---

## Abstract

Posting attractive facial photos is part of everyday life in the social media era. Motivated by the demand, we propose a lightweight method to automatically and efficiently beautify the shapes of both portrait and non-portrait faces in photos, while allowing users to customize the beautification of individual facial features. Previous methods focus on the beautification of mostly frontal and neutral faces, without incorporating user controllability in the beautification process. To address these restrictions, we propose the *Facial Reshaping Operator* representation, which is affine-invariant, captures the pairwise geometric configuration of facial landmarks, and allows for efficient face beautification with the user-specified weights of individual facial parts. We also propose an unsupervised beautification method in the operator space of faces, where an input face is iteratively pulled towards a local nearby density mode with improved attractiveness. Our method distinguishes itself from the commercial beautification tools in that it mildly enhances facial shapes without altering makeups or complexions, which complements these tools that lack fine-grained control on the attractiveness of facial shapes for users. The experimental results show that our method improves facial shape attractiveness for a large range of poses and

---

\*Corresponding author: Hubert P. H. Shum

Email addresses: shanfeng.hu@northumbria.ac.uk (Shanfeng Hu),  
hubert.shum@durham.ac.uk (Hubert P. H. Shum), 07395@buaa.edu.cn (Xiaohui Liang),  
frederick.li@durham.ac.uk (Frederick W. B. Li), nauman.aslam@northumbria.ac.uk  
(Nauman Aslam)

expressions, demonstrating the potential of applicability to photos seen on the social media such as Facebook and Instagram everyday.

*Keywords:* Face Beautification, Facial Attractiveness, Facial Reshaping Operator, Facial Geometry

---

## 1. Introduction

Driven by the culture of social media, the need of appearing attractive (Seidman & Miller, 2013) has stimulated a body of research on facial attractiveness analysis (Zhang et al., 2017). Accordingly, the culture has also demanded the development of commercial systems and tools that allow users to automatically and efficiently enhance their photos before posting on the social media platforms such as Facebook and Instagram. The study of (Hu et al., 2014) shows that nearly half of the photos sampled from Instagram belong to the *Selfies* and *Friends* categories, with both portrait and non-portrait faces being the main subjects presented in them. However, most of the existing commercial tools achieve face beautification by removing darkness and smoothing facial skins (Ciuc et al., 2013), which may not be desirable in some cases when users want to preserve their original makeups and complexions (Dang et al., 2019) while still being able to improve the overall facial attractiveness of the posed photos. This scenario motivates us to propose an intelligent system that only enhances the shapes of faces in the photos, which have significant influences on the attractiveness (Leyvand et al., 2008), without altering the makeups or complexions. We expect our system to complement the existing face beautification tools in providing isolated, fine-grained control on the attractiveness of facial shapes for users.

While facial attractiveness is a subjective notion, some studies have shown that regardless of the ages, genders and races of human observers, their preferences towards more attractive faces share some objective natures (Cunningham et al., 1995; Slater et al., 1998; Winston et al., 2007). This motivates the use of computer techniques for facial attractiveness analysis and enhancement. There

have been studies on the impact of the geometry of faces on attractiveness evaluation (Schmid et al., 2008) and enhancement (Liao et al., 2012). However, their rule-based methods cannot quantify how much a face departs from or conforms to the average of a group of face, which is hypothesized to be attractive (Grammer & Thornhill, 1994). In contrast, data-driven methods that require human-annotated attractiveness scores for face analysis and beautification have been proposed (Eisenthal et al., 2006; Leyvand et al., 2008). More recently, deep learning techniques have been applied to the evaluation and enhancement of facial attractiveness (Gan et al., 2014; Li et al., 2015). The main limitation of these supervised methods is that they require a manually annotated face dataset with attractiveness scores for training, which is labour-intensive to obtain and may bias towards specific facial traits (e.g. races) if the dataset is not sufficiently large-scale. The dependency of these methods on annotations also presents a major hurdle to the beautification of non-frontal, non-neutral faces in most photos, because it remains challenging to rate the attractiveness of a facial shape when the pose or expression variations are confounded (Zhang et al., 2017). The other drawback of these methods is that they require non-linear iterative optimization steps to gradually improve the estimated attractiveness of a given face, which is computationally costly and prone to local optimum. These challenges motivate us to propose a more lightweight method for face beautification, which only requires a collection of unannotated faces while enabling much more efficient face reconstruction with the global optimum.

Firstly, we propose to approach the beautification of non-frontal, non-neutral faces in a geometrically controllable way. Departing from the coordinate-based face representations of (Leyvand et al., 2008; Liao et al., 2012; Chen et al., 2014), we propose to encode the geometry of a face using the orthogonal projection operator onto the subspace of the facial landmarks. Due to the affine-invariance of our operator representation, it frees the face beautification process from the underlying nuisance facial landmarks transformations, such as translation, rotation, and scaling. Furthermore, it fully encodes the pairwise geometric configuration of facial landmarks, which enables users to prescribe different levels

of beautification for individual facial parts. Comparing with the non-linear face reconstruction methods in (Leyvand et al., 2008; Liao et al., 2012; Chen et al., 2014), the reconstruction of facial landmarks from our operator representation  
60 is a linear projection, which is much more efficient and guaranteed to be globally optimal.

On top of our operator representation of faces, we propose to identify beautiful face patterns as the local density modes in the operator space of faces. The idea is that these density modes represent the local clusters of facial shapes and  
65 exhibit stronger tendency of symmetry and averageness (Grammer & Thornhill, 1994). We observe that to obtain high-fidelity face beautification results, users typically want to apply minimum changes to the geometry of an input face while keeping the original pose and expression intact. Therefore, we formulate the beautification process of an input face as pulling it towards a local  
70 nearby density mode, which can be efficiently found using the mean-shift method (Georgescu et al., 2003). As the method successively averages the face operators in a local neighbourhood until convergence, it is locality-sensitive averaging for face beautification. The locality-sensitive averaging allows us to adapt the beautification of a particular face to the local vicinity in the operator space of  
75 faces, so that the original pose and expression variations can be preserved.

Results show that our method improves facial attractiveness for a wide range of non-frontal poses and non-neutral expressions, without relying on any human-annotated attractiveness scores for training (Leyvand et al., 2008; Chen et al., 2014; Li et al., 2015). 70% of the human subjects we interviewed prefer our beau-  
80 tified results for 100 frontal portraits, while 65% of the subjects prefer the beautified results for 100 general facial images. Results also show that our method preserves user-specified facial parts for face beautification. The whole process takes less than 1 second to finish on a laptop and therefore allows for continuous user interactions. Our publicly available source codes can be downloaded  
85 from this link: [https://drive.google.com/open?id=1NonS5WQedtxejTDh-m\\_Ym3MZSk21H54p](https://drive.google.com/open?id=1NonS5WQedtxejTDh-m_Ym3MZSk21H54p)

The contributions of this paper include:

- Theoretically, we propose to represent the geometry of a face using the orthogonal projection operator onto the subspace of the facial landmarks (Chen et al., 2015). This representation is affine-invariant, captures the pairwise geometric configuration among facial landmarks, and allows for efficient face beautification with the user-specified weights of individual facial parts. We also propose to formulate the beautification process of a face as locality-sensitive mode seeking (Li & Tang, 2018) using the mean-shift method. The method is capable of beautifying non-frontal, non-neutral faces by applying minimum changes to input faces via locality-sensitive averaging.
- Practically, we demonstrate, to the best of our knowledge, the first intelligent system that is able to beautify the shapes of both portrait and non-portrait faces in a purely unsupervised manner. The user study results validate that 70% of the examined users prefer our beautification results for portrait faces, and that 65% of them prefer our results for non-portrait faces. Our system does not alter the original facial makeups or complexions of photos, which has the potential of complementing the existing commercial tools that modify these non-shape factors.

The remainder of the paper is structured as follows. We review existing work in Section 2 and describe our face beautification approach in 3. We present our results in Section 4 and conclude the paper in Section 5.

## 2. Related Work

In this section, we briefly review existing work related to facial modeling and editing, facial attractiveness analysis, and facial attractiveness enhancement.

**Facial Modeling and Editing.** One approach to face beautification in natural photos is reconstructing 3D faces from 2D images and applying the 3D face rectification method of (Liao et al., 2012). Despite the development of statistical shape models (Banz & Vetter, 1999; Tena et al., 2011; Maleš et al.,

2019) and example-based models (Kemelmacher-Shlizerman et al., 2011; Hassner, 2013), 3D face reconstruction with a wide range of poses and expressions remains a challenging ill-posed problem. While the work of (Yang et al., 2011) can be used for facial component transfer, it does not address the problem of face  
120 beautification. Therefore, following (Leyvand et al., 2008; Chen et al., 2014), we focus on 2D face beautification in this work.

Textural, expressive, and photometric traits also have important influence on the perception of facial attractiveness. As a result, a large body of research have been devoted to the editing of these traits in 2D images, such as face makeup  
125 (Guo & Sim, 2009; Scherbaum et al., 2011; Zhang et al., 2019), expression editing (Yang et al., 2011, 2012), pimples removal (Brand & Pletscher, 2008), photometry correction (Joshi et al., 2010), and hair decoration (Pasupa et al., 2019). Our work complements these methods in that we focus on editing the geometric trait of faces in 2D images.

130 **Facial Attractiveness Analysis.** Computer techniques have been used for facial attractiveness analysis over two decades. We refer the readers to (Laurentini & Bottino, 2014) for an excellent review of the field. (Grammer & Thornhill, 1994) composed different facial images to verify the effect of symmetry and averageness on the perception of facial attractiveness. Later, (Zhang  
135 et al., 2011) validated the effect of averageness using geometrically transformed faces. (Schmid et al., 2008) took a rule-based approach to examine that human-annotated attractiveness scores were consistent with that predicted by neoclassical canons, symmetries, and golden ratios. (Eisenthal et al., 2006) represented the first to use machine learning for facial attractiveness prediction. They employed human subjects to rate a library of faces and trained attractiveness regression models using the appearance and geometry features of faces. Following  
140 this, a number of learning-based facial attractiveness prediction methods have been proposed (Zhang et al., 2017; Gan et al., 2014; Chen et al., 2014). As it is ambiguous to rate the attractiveness of non-frontal, non-neutral faces, previous  
145 work mostly focuses on the analysis of frontal portraits. Therefore, we propose an unsupervised beautification method for non-frontal, non-neutral faces.

**Facial Attractiveness Enhancement.** Compared with facial attractiveness analysis, the enhancement problem has received relatively less research attention. (Liao et al., 2012) optimized the shape of a 3D face model by de-  
150 forming it towards the beauty cannons summarized by (Schmid et al., 2008). The method works well for neutral faces but has inherent difficulties of generalizing to non-neutral faces. (Leyvand et al., 2008) was the first to approach attractiveness enhancement using machine learning. Their results validated the feasibility of a data-driven approach to face beautification. Recently, (Chen  
155 et al., 2014) searched for a convex combination of attractive faces while maximizing its resemblance to the original face. However, both methods require human annotations of attractiveness scores for training and cannot work for non-frontal, non-neutral faces, which are actually the main subjects of real-world photos. Deep learning has also been applied to facial attractiveness en-  
160 hancement (Li et al., 2015). However, the resulting system does not allow for the controllability of beautification due to the black-box nature of deep learning techniques. In contrast from the methods of (Leyvand et al., 2008; Chen et al., 2014; Li et al., 2015), our proposed one works for faces with non-frontal poses and non-neutral expressions in images. Benefiting from the linearity of our op-  
165 erator representation, our method easily allows users to preserve certain facial parts while beautifying the remaining.

### 3. Our Facial Operator Approach to Face Beautification

Fig. 1 provides an overview of our face beautification method. Given an input facial image, we first detect (a) a collection of facial landmark points  
170 along the jawline and the contours of the eyebrows, eyes, nose, and mouth. We then compute (b) an operator to capture the input facial shape, which is the orthogonal projection matrix onto the subspace of the detected landmarks. After that, we perform (c) face beautification on the operator by pulling it towards a local nearby density mode in the operator space of faces. We then  
175 reshape (d) the original landmarks using the optimized operator so that the



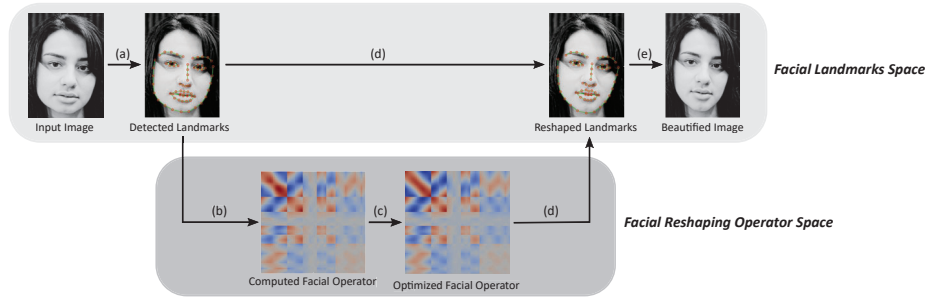


Figure 1: **The Overview of Our Face Beautification Method.** The computation steps from (a) to (e) are facial landmarks detection, facial operator computation, facial operator optimization, facial landmarks reconstruction, and facial image warping respectively.

geometric configuration of the input face can be enhanced. Finally, we warp (e) the input image using the original and the reshaped landmarks to produce an beautified image. We also describe our system using the pseudocode in

Algorithm 1. We describe the steps in the following sections.

---

**Algorithm 1:** The overall process of our face beautification system

---

**Input:** A facial image  $I$ ;

A vector of beautification weights  $\mathbf{w} \in \mathbb{R}^n$ , where  $n$  is the number of facial landmarks and  $0 \leq \mathbf{w}_i \leq 1$  is the strength of modification to the  $i$ -th landmark;

A dataset of unannotated face images.

**Output:** A user-controlled beautified facial image  $I^*$ .

- 1 Detect a collection of facial landmarks  $P \in \mathbb{R}^{n \times 3}$  from  $I$  (Section 3.1), where the homogeneous coordinate of the  $i$ -th landmark is  
 $P_i = (P_{i1}, P_{i2}, 1)$ ;
  - 2 Compute the facial operator representation of  $P$  as  $M = P(P^T P)^{-1} P^T$  (Equation 2, Section 3.2), which is the orthogonal projection operator onto the subspace of  $P$ ;
  - 3 Compute an enhanced facial operator  $M^*$  from  $M$  by repeatedly shifting it towards a local nearby density mode of the given face dataset until convergence (Equation 5, Section 3.3);
  - 4 Reshape the original facial landmarks  $P$  into the enhanced ones  
 $P^* = [(1 - W)M + WM^*]P$  (Equation 9, Section 3.4), where  
 $W = \mathbf{w}\mathbf{w}^T$  is the matrix of pairwise beautification weights;
  - 5 Warp the original facial image  $I$  into the beautified one  $I^*$  using the original and enhanced facial landmarks  $P$  and  $P^*$  (Section 3.5);
- 

### 3.1. Facial Landmarks Detection

Given an input facial image  $I$ , the first step of our approach is to detect a collection of facial landmarks  $P \in \mathbb{R}^{n \times 3}$  for representing the geometry of the input face, where  $n$  is the number of landmark points. The coordinates of the  $i$ -th landmark is  $P_i = (P_{i1}, P_{i2}, 1)$ , where  $P_{i1}$  and  $P_{i2}$  are the  $x$  and  $y$  coordinates on the image plane respectively. The last component is the homogeneous coordinate 1, which we need to derive our facial operator representation of  $P$  in Section 3.2. We adopt the method of (Le et al., 2012) for detecting a num-

ber of  $n = 68$  facial landmark points along the jawline and the contours of the  
 190 eyebrows, eyes, nose and mouth. This method builds a local shape model for  
 each individual facial part, which is capable of handling a wide range of facial  
 expression and pose variations.

### 3.2. Facial Operator Representation

After detecting a collection of facial landmarks  $P$  from an input image  $I$ ,  
 195 we aim at deriving a more effective representation of  $P$  for facial shape ma-  
 nipulation. Our idea is to represent  $P$  as the orthogonal projection operator  
 $M \in \mathbb{R}^{n \times n}$  onto the coordinate subspace of  $P$ . Traditionally, the normalized  
 coordinates of  $P$  and the edge lengths of the delaunay triangulation have been  
 widely used (Leyvand et al., 2008; Chen et al., 2014). However, these low-level  
 200 representations cannot be effectively converted back to the coordinate space,  
 making the facial reshaping process non-linear with potentially bad local min-  
 imas. In contrast, our operator representation  $M$  is naturally invariant to the  
 nuisance affine perturbations of the detected facial landmarks, which geomet-  
 rically correlates every pair of landmarks in a human-understandable way. In  
 205 Section 3.4, we show that it significantly simplifies the facial reshaping process to  
 linear projection, allowing for flexible user controllability on the beautification  
 process

**Formulation.** To solve for the operator representation  $M$  that geometri-  
 cally correlates every pair of facial landmarks, we consider representing each  
 landmark  $P_i$  as the linear combination of all the landmarks. Because the num-  
 ber of landmarks  $n$  is normally greater than the coordinate dimension (i.e. 3),  
 we further minimize the L2-norm of the combination coefficients so that the rep-  
 resentation can be uniquely determined. This leads to the following constrained  
 minimization problem for each landmark:

$$\mathbf{x}^* = \arg \min \|\mathbf{x}\|_2^2, \text{ subject to } P_i = \sum_{k=1}^n \mathbf{x}_k P_k \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^n$  is a vector of linear combination coefficients for each landmark  
 and  $\|\mathbf{x}\|_2^2 = \sum_{k=1}^n \mathbf{x}_k^2$  is the L2-norm of  $\mathbf{x}$  to be minimized. Solving (1) for each

landmark individually gives us the matrix representation  $M$ :

$$\overbrace{M = P(P^T P)^{-1} P^T}^{\text{Facial Reshaping Operator}} \quad (2)$$

where  $T$  is the matrix transpose operator and each row of  $M$  is the optimal solution to (1) for the corresponding landmark. We can verify from (2) that the representation is symmetric ( $M = M^T$ ), idempotent ( $M = M^2$ ), and exact ( $P = MP$ ), confirming that it is the orthogonal projection operator onto the subspace of the facial landmarks. This construction fundamentally changes the formulation of facial shape manipulation from the traditional non-linear coordinates optimization to our linear projection.

**Affine-Invariance.** Here, we verify that our operator representation  $M$  is guaranteed to be affine-invariant, which allows for efficient facial shape analysis and manipulation without nuisance facial landmark coordinate transformations involved in the process. For each landmark  $P_i$ , it can be seen from (1) that the constraint still holds when we apply any linear transformation  $L \in \mathbb{R}^{3 \times 3}$  to all the landmarks,  $LP_i = \sum_{k=1}^n \mathbf{x}_k(LP_k)$ . This shows that the optimal combination coefficients (i.e. rows of  $M$ ) remain the same under any linear (rotation, uniform scaling, and shearing) perturbation in real-world facial images. By applying any homogeneous translation  $T = (x, y, 1) \in \mathbb{R}^3$  to all the landmarks, we can further obtain  $P_i + T = \sum_{k=1}^n \mathbf{x}_k(P_k + T) = \sum_{k=1}^n \mathbf{x}_k P_k + (\sum_{k=1}^n \mathbf{x}_k)T$ . Because the last coordinate components of all the landmarks are 1, we have  $\sum_{k=1}^n \mathbf{x}_k = 1$  from the constraint in (1). Therefore, the equation  $P_i + T = \sum_{k=1}^n \mathbf{x}_k(P_k + T)$  holds, which confirms that the combination coefficients are also invariant to any translation of the facial landmarks.

**Geometric Significance.** On top of the affine-invariance, our operator representation  $M$  also has clear geometric significance that is not provided by deep learning methods (Gan et al., 2014; Li et al., 2015). This allows it to fully capture the geometric configurations for every pair of facial landmarks, which enables part-based user control in face beautification. We reveal this by finding

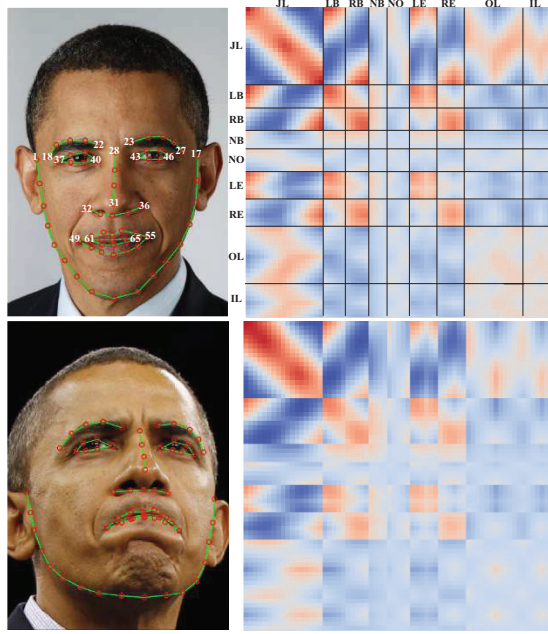


Figure 2: **The Visualization of Our Operator Representation of Facial Landmarks.**

*Left:* two faces with different expressions and poses, with the corresponding facial landmarks rendered on top of the faces; the white numbers on the top left show the indices of the 68 facial landmarks. *Right:* the rendered images of the two corresponding operator representations, with blue and red colors representing low and high values respectively; the black lines separate different facial parts for clearer visualization. These parts include the jawline (**JL**), left eyebrow (**LB**), right eyebrow (**RB**), nose bridge (**NB**), nostril (**NO**), left eye (**LE**), right eye (**RE**), outer lip (**OL**), and inner lip (**IL**). The pairwise geometric configurations represented by our operator enables part-based user control in face beautification.

the analytical form of each element of  $M$  as follows:

$$M_{ij} = \frac{(P_i - \bar{P})^T S^{-1} (P_j - \bar{P})}{n} + \frac{1}{n} \quad (3)$$

where  $\bar{P} = \frac{1}{n} \sum_{k=1}^n P_k$  is the centroid of the face and  $S = \frac{1}{n} \sum_{k=1}^n (P_k - \bar{P})(P_k - \bar{P})^T$  is the covariance matrix of the landmark coordinates. Now, it becomes clear that the diagonal elements of  $M$  depend on the normalized squared Euclidean distance from each landmark to the face centroid, while the non-diagonal elements depend on the dot product between the pair of normalized vectors from the two corresponding landmarks to the centroid.

**Facial Operator Visualizations.** To demonstrate the geometric significance of our operator representation, we visualize the operators of two faces with different expressions and poses in Fig. 2. It can be seen that each facial part has its a distinct pattern within the matrix. These patterns come from the configurations of the facial parts relative to each other, which vary according to the facial shape. As the distances from the landmarks to the face centers are encoded in the diagonals of the operators, the elements corresponding to the outermost contours (i.e. the jawline and eyebrows) have relatively larger values. The inner facial parts such as the lips and nose have relatively smaller values. The main difference between the two operators are on the rows and columns corresponding to the jawline and lips, due to the more significant changes caused by expression and pose. It is also possible to infer the symmetry information of facial parts from the visualized operators. Taking the first one for example, the submatrix occupied by the jawline and eyebrows is nearly symmetric about its main diagonal, signifying the vertical symmetries of the two facial parts. In comparison, the symmetry patterns of the second face have been considerably weakened by the changing expression and pose.

### 3.3. Facial Operator Optimization

After computing the operator  $M$  from an input set of facial landmarks  $P$ , the next step of our approach is to optimize  $M$  into  $M^*$  with improved facial attractiveness. The main idea of our method is to iteratively pull  $M$  towards

a local nearby density mode in the operator space of faces. This allows us to preserve the original pose and expression during beautification, while gradually improving the attractiveness of the represented face.

We require a dataset of un-annotated facial operators  $\mathcal{M} = \{M^i\}_{i=1}^m$  for training, where  $m$  is the number of faces and each operator  $M^i \in \mathbb{R}^{n \times n}$  is computed from the corresponding face. While the methods of (Leyvand et al., 2008; Chen et al., 2014; Li et al., 2015) focus on the beautification of frontal and neutral faces, our method can enhance the shapes of faces with non-frontal poses and non-neutral expressions, without requiring human-annotated attractiveness scores for training.

**Feature Transformation.** To preserve the major features of faces in beautification, we propose to transform the original operator representation using the the Principal Component Analysis (PCA) for dimension reduction and selection. By preserving 95% of the dataset variance, we obtain a low-dimensional close approximation of an input operator  $M$  as follows:

$$M = \overline{M} + \sum_{c=1}^C \lambda_c \Gamma_c \quad (4)$$

where  $\overline{M} \in \mathbb{R}^{n \times n}$  is the mean,  $\{\Gamma_c\}_{c=1}^C$  are the principal components in the descending order of the associated variance, and  $\{\lambda_c\}_{c=1}^C$  are the projection coefficients of  $M$  onto these components. Therefore, the operator can be approximated using the mean operator and a set of mutually uncorrelated details. These details are sorted in the descending order of the corresponding principal eigenvalues (i.e. data variances).

As the leading coefficients explain more rapidly changing expressions and poses in the dataset, we choose to use the non-leading coefficients for facial shape optimization. Empirically, we exclude the first two coefficients and subject the remaining  $\{\lambda_c\}_{c=3}^C$  to optimization. We reconstruct the optimized operator as  $M^* = \overline{M} + \sum_{c=1}^2 \lambda_c \Gamma_c + \sum_{c=3}^C \lambda_c^* \Gamma_c$ , where  $\{\lambda_c^*\}_{c=3}^C$  are the coefficients after face optimization. This is visualized in Fig. 3. In the following, we use  $M$  to abbreviate the coefficient representations  $\{\lambda_c\}_{c=3}^C$  for explaining our face beautification method.

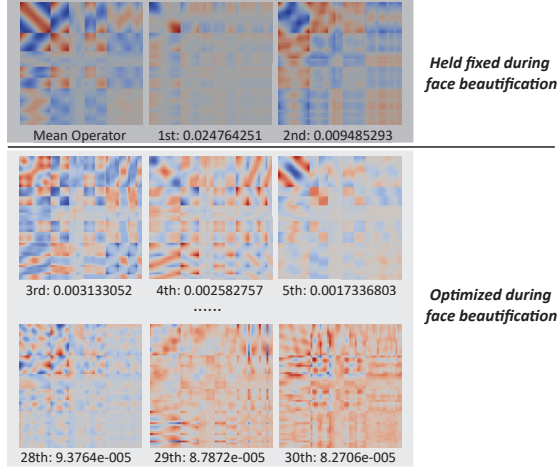


Figure 3: **The Visualization of Our PCA Feature Transformation.** The images show the mean facial operator and the discovered PCA components as sorted by their associated eigenvalues (displayed on the bottom of each small image). We fix the PCA coefficients corresponding to the first and the second components during face beautification, because they encode large-scale pose and expression variations and modifying them leads to noticeable facial image distortions. We only optimize the remaining PCA coefficients that mostly encode intrinsic facial shape features.

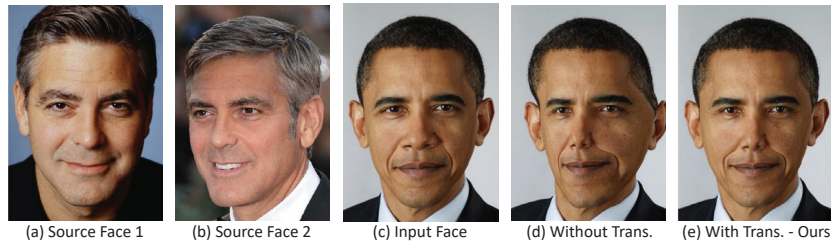


Figure 4: **The Effect of Feature Transformation on Facial Shape Interpolation.** (a): a frontal and neutral source face. (b): another source face with pose and expression. (c): an input face. (d): the interpolated face generated by applying the mean of the two source facial operators to the input face without feature transformation. (e): the interpolated face generated using our feature transformation method.





Figure 5: **The Effect of Feature Transformation on Face Beautification.** (a): an input face. (b): the beautified face without using feature transformation. (c): the beautified face with our feature transformation method.

280 We show an example of face interpolation in Fig. 4, where we apply the mean of the facial operators of two source faces to an input face with and without feature transformation. We evaluate this task because the averaging operation is the building block of our unsupervised face beautification method. It can be seen that while the two source faces have different expressions and poses, they share the consistent intrinsic facial geometric configuration. However, when  
285 their shapes are naively interpolated and transferred to a target face, some extrinsic factors creep in and lead to undesirable face distortions. By excluding the leading PCA coefficients, it can be seen that the transferred face has no noticeable distortions. Furthermore, we show an example of face beautification with and without feature transformation in Fig. 5. It can be seen that the  
290 beautified image with feature transformation is much more realistic compared with that without feature transformation. The output images in Fig. 4 and 5 are generated using the facial image warping method in Section 3.5.

**Locality-sensitive Face Beautification.** Motivated by the averageness  
295 property of facial attractiveness (Grammer & Thornhill, 1994; Schmid et al., 2008; Zhang et al., 2011), we propose to optimize an input facial operator  $M$  by iteratively averaging its local nearby samples in the operator space of faces. Our idea is to regard faces at the local cluster centers (i.e. local average faces) as the locally most attractive, which to our knowledge is first proposed in this work.  
300 Different from the methods of (Grammer & Thornhill, 1994; Schmid et al., 2008;

Zhang et al., 2011), we ensure the locality of the averaging operation so that the averaged faces have improved attractiveness while remaining similar to the input. According to the mean-shift formulation of (Comaniciu & Meer, 2002), this iterative locality-sensitive averaging operation is guaranteed to converge to  
305 a local nearby density mode in the operator space, under mild assumptions of the kernel function used for measuring the distance between any pair of facial operators.

We formulate the idea of iterative local averaging and obtain our face beautification formula at one scale as follows:

$$\overbrace{\Psi(M) = \frac{\sum_{i=1}^m [h_i^{-C} e^{-||M^i - M||^2 / 2h_i^2}] M^i}{\sum_{i=1}^m [h_i^{-C} e^{-||M^i - M||^2 / 2h_i^2}]}}^{\text{Locality-sensitive Face Averaging}} \quad (5)$$

where  $M$  is an input facial operator using our PCA feature representation, each  $M^i$  is a training sample from the face dataset  $\mathcal{M}$ ,  $m$  is the number of training  
310 samples in  $\mathcal{M}$ , and  $C$  is the dimension of the PCA representation. It can be seen from (5) that the distance between  $M$  and  $M^i$  is measured as the Gaussian kernel function  $e^{-||M^i - M||^2 / 2h_i^2}$ , which assigns higher weights to geometrically more similar faces and lower weights to faces that are farther away from the input. Compared with the equal or linear weighting kernels, the exponential  
315 locality of the Gaussian kernel function allows for better preservation of the input pose and expression during beautification.

To account for varying sample densities, we associate an adaptive scale  $h_i$  with each face in the dataset:

$$h_i = ||M^i - M^{\gamma m}||^2 \quad (6)$$

where  $\gamma \in (0, 1)$  and  $M^{\gamma m}$  is the  $\gamma m$ -nearest neighbor to  $M^i$  in the dataset. When  $\gamma$  is set to a small value, only very close samples can contribute to the averaging and thus only small-scale facial features will be modified. When  $\gamma$  is  
320 set to a larger value, more distant samples will be considered and larger-scale facial features will be enhanced. Empirically, we choose the set of scales  $\gamma \in \{0.005, 0.01, 0.02, 0.04, 0.08\}$  because they are able to cover both local nearby

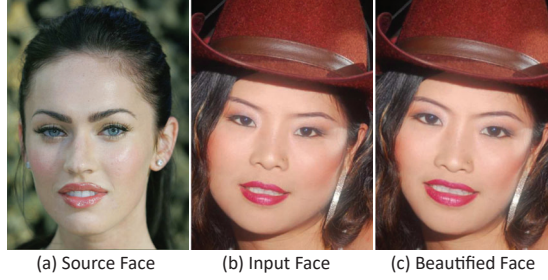


Figure 6: **The Example of Face Beautification via Shape Transfer.** (a): a source face. (b): an input face to be beautified. (c): the beautified face with the shape coming from the example. The shape transfer is efficiently done by applying the operator of the source face to the input as linear projection.

and more distant training samples for facial shape optimization. For each  $\gamma$ , we compute the corresponding adaptive scale  $h_i$  for each training sample  $M^i$  and iteratively evaluate the locality-sensitive averaging operation in (5) until  
 325 convergence. The final optimized operator  $M^*$  is computed as the mean of the convergent solutions of all scales.

### 3.4. Facial Landmarks Reshaping

With the original facial landmarks  $P$  and the optimized facial operator  $M^*$   
 330 computed by our locality-sensitive averaging method, the next step of our approach is to find a set of new landmarks  $P^*$  that are consistent with the enhanced facial shape represented by  $M^*$ . We show that our operator representation reduces the process to a linear projection, which is significantly less costly compared with the non-linear optimization method of (Leyvand et al., 2008; Chen  
 335 et al., 2014). Besides the global optimality of our method, it allows users to flexibly control the beautification weight of each individual facial part, thereby achieving user-satisfied beautification results.

**Formulation.** To ensure the quality of face beautification, we minimize the distance between the original and the new landmarks while enforcing the consistency of the new landmarks with the optimized operator:

$$P^* = \arg \min \|X - P\|_F^2, \text{ subject to } X = M^* X \quad (7)$$

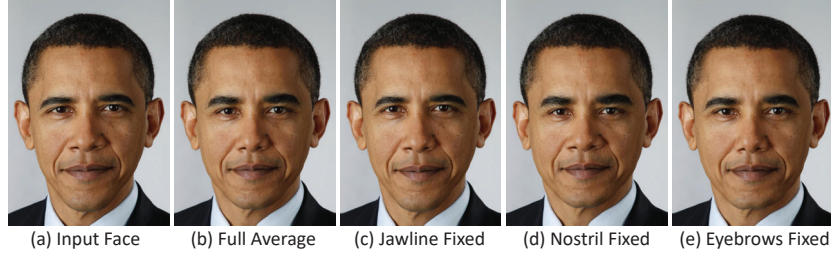


Figure 7: **The Example of User-controlled Face Averaging.** (a): an input face. (b): a full average face created by applying the mean facial operator to the input. (c): the average face with the original jawline shape by setting the corresponding beautification weight to 0. (d): with the original nostril shape. (e): with the original left and right eyebrow shapes.

where  $X \in \mathbb{R}^{n \times 3}$  is a set of new landmarks to be found and  $\|X - P\|_F^2 = \sum_{i=1}^n \sum_{j=1}^2 (X_{ij} - P_{ij})^2$  measures the deviation of the two sets of landmarks. The constraint enforces that the new landmarks should be within the subspace corresponding to the operator  $M^*$ , thereby reconstructing the encoded pairwise geometric configurations of the facial landmarks. The optimal solution to (7) is the linear projection of the original facial landmarks onto the new optimized operator:

$$\overbrace{P^* = M^* P}^{\text{Facial Reshaping}} \quad (8)$$

Essentially,  $M^*$  acts as the facial reshaping operator that enhances an input face via linear projection, which is significantly more efficient than the non-linear optimization process of (Leyvand et al., 2008; Chen et al., 2014).

Fig. 6 shows an example of face beautification via shape transfer. It can be seen that the modified face appears to resemble the source face as indicated by the thinner jawline, longer eyebrows, bigger eyes and mouth. This is efficiently done by linearly projecting the input facial landmarks using the operator of the source face.

**User Controllability.** Now, we consider the problem of user controllability in face beautification. Sometimes, users may want to prescribe different levels of beautification for individual facial features, such as preserving the shapes of the eyebrows more while enhancing the shapes of the jawline more. This

can be naturally done using our operator representation. We denote a vector of beautification weights as  $\mathbf{w} \in \mathbb{R}^n$ , where  $0 \leq \mathbf{w}_i \leq 1$  is a user-specified value controlling how much the facial shape related to the  $i$ -th landmark can be altered. We group the landmarks into the *jawline*, *left eyebrow*, *right eyebrow*, *nose bridge*, *nostril*, *left eye*, *right eye*, *outer lip*, and *inner lip*, as visualized in Fig. 2 and described in Appendix 1. The landmark weights within each group are equal. As a result, the beautification weight matrix  $W = \mathbf{w}\mathbf{w}^T \in \mathbb{R}^{n \times n}$  indicates the weight configuration of the whole face:  $W_{ij}$  is larger if the relative geometric configuration of the landmarks  $i$  and  $j$  should be enhanced more, and  $W_{ij}$  is smaller if the configuration needs to be preserved more. We compute the reshaped facial landmarks as follows:

$$\overbrace{P^* = [(1 - W)M + WM^*]P}^{\text{User-controlled Facial Reshaping}} \quad (9)$$

where  $M$  and  $M^*$  are the original and the optimized facial operators respectively. If  $\mathbf{w}$  is a vector of 0, no beautification is applied and the original face is recovered. If  $\mathbf{w}$  is a vector of 1, the new face is generated without preserving any features of the original. By adjusting the weights of the facial parts, users can enhance  
350 input faces while preserving certain desired features.

Fig. 7 shows an example of applying the mean facial operator to reshape an input face, while keeping some of the original facial parts fixed during reshaping. The mean operator is computed on the training set of (Le et al., 2012) and the created full average face appears to have more round and symmetric facial  
355 features than the input. By setting the beautification weights of the facial landmarks on the jawline to 0, the partial average face looks more similar to the input than the full average, as the jawline has a global influence on the shape of the whole face. By fixing the nostril or the eyebrows instead, the results are more similar to the full average face, but the shapes of the nostril  
360 and eyebrows still resemble the originals. Our user-controlled facial reshaping method produces satisfactory customized results in all cases.

### 3.5. Facial Image Warping

Finally, given an input facial image  $I$  and the detected facial landmarks  $P$ , the final step of our approach is to produce an beautified image  $I^*$  using the optimized facial landmarks  $P^*$ . Our goal is to geometrically deform the input image so that the original facial landmarks can be matched to the optimized ones, while ensuring that the deformed image remains realistic and high-quality. To this end, we adopt the Moving Least Squares method of (Schaefer et al., 2006) for fitting an rigid transformation on each image location. To more efficiently process high-resolution facial images that are popular nowadays, we modify the original method by sampling a low-resolution grid (10% of the image resolution) on the input image and deforming the grid using the original and the optimized landmarks. Afterwards, for each quad on the deformed grid, we compute a backward perspective transformation that maps the quad to the corresponding one on the original grid. Finally, we use these backward transformations to fill the output image using the corresponding pixels from the input image. Our modified method is capable of producing a high-quality, high-resolution beautified image within a second. We use it to generate all of the modified face images in this work.

## 4. Results

### 4.1. Full-face Beautification Results

We show our full-face beautification results in Fig. 8. We provide more results on general facial images in Appendix 2 and results on frontal portraits in Appendix 3. We generate these results using the popular Helen facial image dataset (Le et al., 2012), which consists of 11,147 high-resolution facial images for training and 2,330 images for testing. These images cover a diverse range of genders, races, ages, poses, and expressions. The diversity of the Helen dataset is important for well representing the local density modes in the face space, which our method seeks for face beautification. Despite the diverse backgrounds, shapes, poses, and expressions of the input faces, our method manages

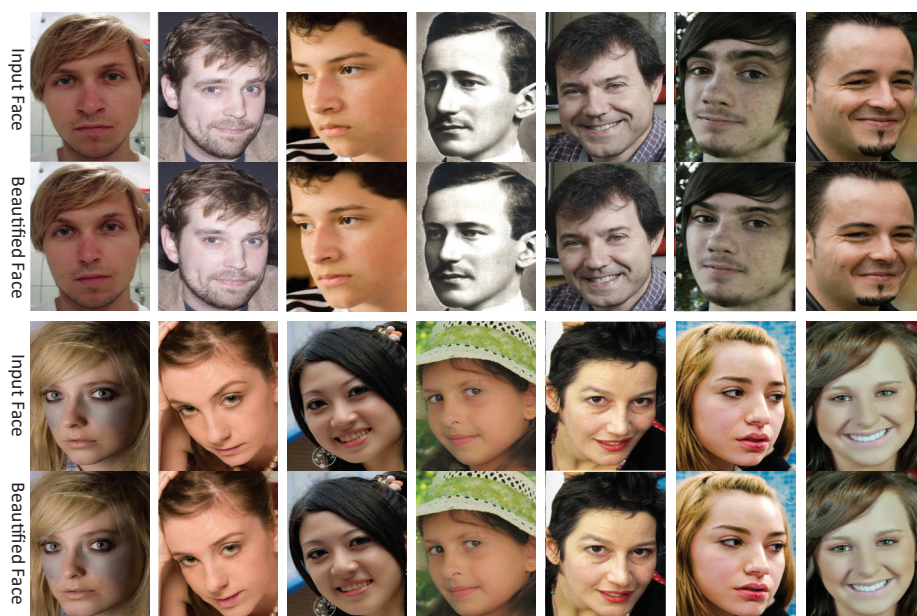


Figure 8: **Our Full-face Beautification Results.** These faces are selected to cover a wide range of gender (male and female), pose (frontal and non-frontal), and expression (neutral and non-neutral). Our method is purely unsupervised and does not require any human-annotated facial attractiveness scores for training.



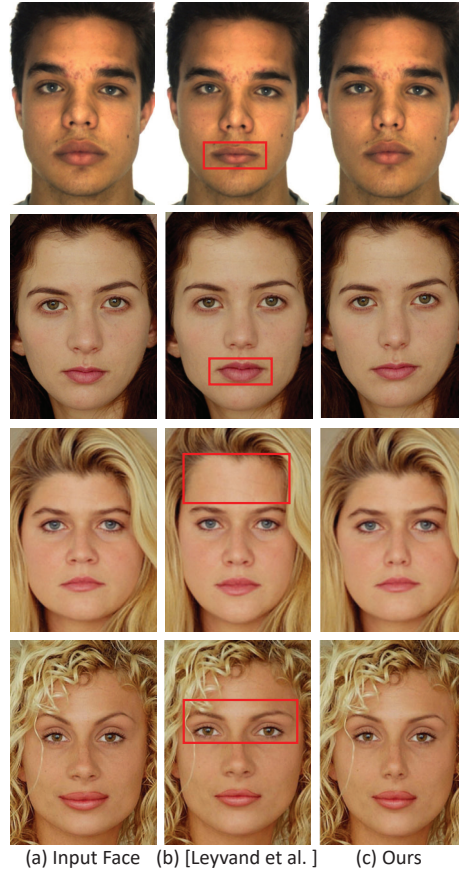


Figure 9: **The Comparison of Our Unsupervised Method with the Supervised Method of (Leyvand et al., 2008).** (*a*): the input facial images. (*b*): the beautified images taken from (Leyvand et al., 2008). (*c*): the beautified images generated by our method. We use the red boxes to highlight the overly modified characteristic facial parts that some end-users may want to preserve during beautification.





Figure 10: **The Comparison of Our Unsupervised Method with the Supervised Method of (Chen et al., 2014).** *Left:* the input facial images. *Middle:* the beautified images taken from (Chen et al., 2014). *Right:* the beautified images generated by our method. We use the red boxes to highlight the overly modified characteristic facial parts that some end-users may want to preserve during beautification.

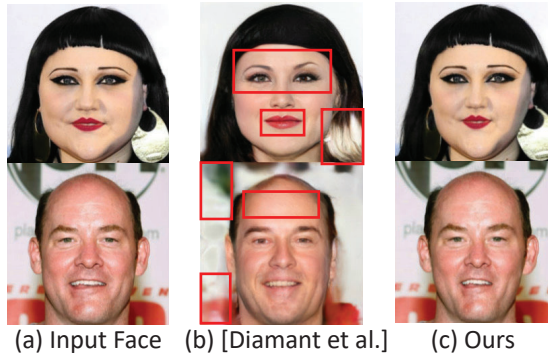


Figure 11: **The Comparison of Our Unsupervised Method with the Semi-supervised Method of (Diamant et al., 2019).** *Left:* the input facial images. *Middle:* the beautified images taken from (Diamant et al., 2019). *Right:* the beautified images generated by our method. We use the red boxes to highlight the overly modified characteristic image regions that some end-users may want to preserve during beautification.

to automatically optimize the input facial geometric configurations and produce high-resolution images with improved facial attractiveness. Importantly, the poses and expressions of the input faces are well preserved by our method. Note that neither human-annotated attractiveness scores (Leyvand et al., 2008; Chen et al., 2014; Li et al., 2015) nor 3D face modeling (Liao et al., 2012) are required by our method. Our face beautification results are all generated in a purely unsupervised manner.

**Comparisons with Supervised Methods.** Here, we compare our method with that of (Leyvand et al., 2008) and (Chen et al., 2014). These two methods only work on frontal portraits and therefore cannot generalize well to non-frontal poses and non-neutral expressions. Besides, they both require human-annotated facial attractiveness scores for training, which unfortunately are difficult to obtain for non-frontal and non-neutral faces that exist in many real-world applications.

Due to the limited number of results provided in (Leyvand et al., 2008) and (Chen et al., 2014), we can only compare with a few facial images taken from the original papers in Fig. 9 and 10. Both our unsupervised method

and the two supervised methods improve the perceived attractiveness of the input faces. While the supervised methods are shown to apply more intense  
 410 modifications to the eyebrows and eyes, our method preserves more features of these regions. The beautified faces generated by our method appear more similar to the input, which is because our method performs beautification via iterative locality-sensitive averaging. Therefore, it is capable of maintaining some of the input facial features that are normally missing during the supervised  
 415 optimization process of (Leyvand et al., 2008) and (Chen et al., 2014).

**Comparisons with Semi-supervised Methods.** Here, we compare our method with that of (Diamant et al., 2019), which is the most recent method that uses facial attractiveness annotations to train a beauty-conditional generative adversarial network (GAN) for semi-supervised face beautification. This  
 420 method represents the state-of-the-art research on leveraging deep learning and generative modelling for face beautification. We take the mildly beautified photos from the original paper and compare our results with them in Fig. 11. It can be clearly seen that the method of (Diamant et al., 2019) modifies mainly the facial textures, resulting in smoother skins (the top two faces on the middle  
 425 column) and a new facial identity (the bottom face on the middle column). In contrast, our method solely enhances the facial shapes (e.g., the thinner jawline as shown on the bottom right), without altering the original textures or complexions. As a result, our method can maintain the original facial characteristics while still improving the perceived facial attractiveness.

#### 430 4.2. User Study

To empirically evaluate the effectiveness of our face beautification method, we designed a user study for collecting human preferences towards the original faces or the beautified versions generated by our method. The study consists of two courses, one based on 100 general facial images from the challenging  
 435 Helen test set (Le et al., 2012) and the other based on 100 more controlled frontal portraits we collect from the Internet. We conducted the study on both sets of images to demonstrate the robustness of our method in handling non-

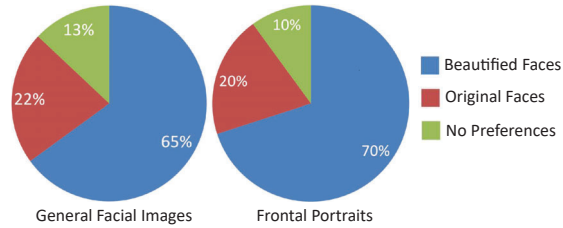


Figure 12: **The User Study of Our Method.** *Left:* the percentages of human subjects preferring the original faces, the beautified faces, or without preferences on general facial images. *Right:* the percentages of human preferences on input faces with frontal poses and neutral expressions.

frontal, non-neutral faces in the wild. We generated the beautified results for both sets of images and presented them in Appendix 2 and 3 respectively. We recruited 12 male and 12 female human subjects aged 22-35. In each course, we randomly presented pairs of the original and the beautified images to each of the subjects, with the order of the images in each pair being randomized. We asked each subject to pick the more attractive face in each pair or to express no preferences. When finished, we obtained 2,347 and 2,085 choices for the two courses respectively. We did not have the full 2,400 choices for either of the two courses because some participants accidentally skipped a small fraction of them without submitting any preferences. Nevertheless, we still had around 97.8% and 86.9% response rates, which were sufficient for our analysis.

As shown in Fig. 12, the subjects in both courses prefer the full-face beautified images over the originals. Remarkably, the percentages of subjects preferring our results on the much more challenging general facial image course, which includes non-frontal and non-neutral faces, are comparable to that on the frontal portraits (70% versus 65%). This demonstrates the robustness of our method for enhancing non-frontal and non-neutral faces, which are very common in many real-world applications. To our knowledge, this is the first time in the field that a method has been shown to be capable of beautifying such challenging faces.

Here, we provide a more detailed analysis of the user study results on the

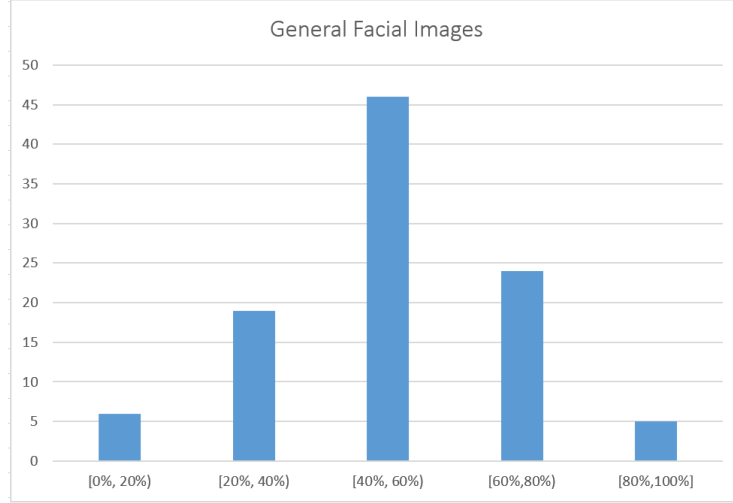


Figure 13: **The Distribution of General Facial Images in the User Study.** For each of the 100 general facial images in the user study, the percentage of the 24 subjects who prefer the beautified result is computed and aggregated here. To be exact, 6, 19, 46, 24, and 5 images receive preference ratings from a percentage of subjects as shown for each bin.

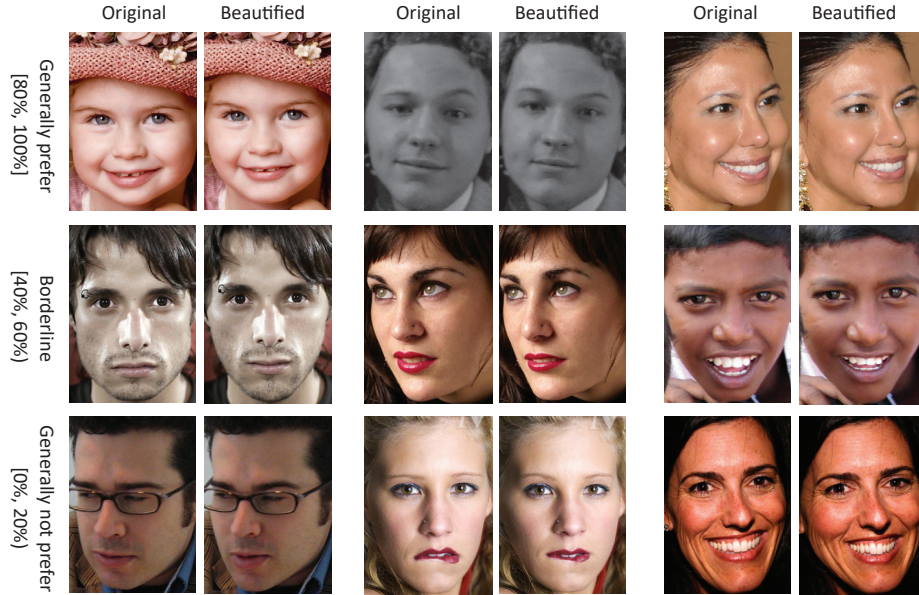


Figure 14: The examples of general facial images that receive preference ratings from a majority (i.e. over 80%) of, a mixed pool (i.e. between 40% and 60%) of, and a minority (i.e. under 20%) of the 24 subjects in the user study.

course of general facial images, because they represent more realistic face variations and are the main targets of this work. As can be seen in Fig. 13, very few of the beautified images are either preferred (i.e. over 80%) or not preferred (i.e. under 20%) by a majority of the 24 subjects. This suggests that achieving consensus on face beautification results among the subjects is very unlikely, which is in line with that around 13% of them have no preferences expressed as shown on the left of Fig. 12.

We show some of these particular images along with those that receive borderline ratings (i.e. between 40% and 60%) in Fig. 14. It can be noticed that for the lower-left pair in the figure, the beautified image has distortions around the frame of the glasses, which is probably why it is generally not preferred by the subjects. As for the lower-middle pair with a very novel mouth configuration, the woman’s eyes become less symmetric after beautification and are therefore not considered as attractive by the subjects. In contrast, the beautified images on the upper row are preferred by a high percentage of the subjects, as they display more symmetric facial features and thinner jaw-lines without introducing noticeable distortions. It is even more interesting to note that the borderline images on the middle row do not show noticeable flaws either, but the beautified versions are not considered as significantly more attractive than the originals by the subjects. Taking the middle left pair as an example, the beautified facial image of the man appears less muscular and as a result the subjects have mixed ratings on it.

#### 4.3. User-controlled Beautification Results

Beyond full-face beautification without user intervention, our method allows users to prescribe a beautification weight in  $[0, 1]$  for each individual facial part. The smaller the weight, the better preserved the shape of the corresponding part. Fig. 15 shows some user-customized face beautification results, for which we allow users to preserve one or more facial parts while fully beautifying the remaining. As shown on the first row, our method generates natural and smooth transitions between the shape of the input mouth and that of the fully beautified





Figure 15: **Our User-controlled Face Beautification Results.** Here, we allow users to preserve one or more facial parts while fully beautifying the remaining. On each row, we show the input facial image on the first column and the full-face beautified images on the fourth column. We show the intermediate beautified images on the second and third columns, with the user-customized beautification weights of the facial parts being preserved. The beautification weights of the parts not being preserved are set to the largest value of 1.0.

version, when users change the beautification weights of the outer and inner lips  
490 from 0.0 (i.e. no beautification) to 0.4, 0.8, and 1.0 (full beautification). Our  
method works equally well for allowing users to control the beautification level  
of the jawline (on the second row) and eyes (on the fourth row) of an input  
face. More challenging is the case when users want to preserve more than one  
facial part, as shown on the third row of Fig. 15. Still, our method generates  
495 intermediate beautified images that naturally interpolate the original and the  
fully enhanced eyebrows and mouth.

#### 4.4. Computational Cost

We list the run time of each computation step of our face beautification  
method in Table 1. The time is measured for processing a typical facial image  
500 of resolution  $1280 \times 960$  on a laptop with an Intel(R) Core i7-6500U 2.5GHZ  
CPU and a 8GB RAM. Overall, the whole beautification process takes less than  
1 second to finish, which allows for continuous user interactions with low latency.  
The steps of facial operator computation (b), facial operator beautification (c),  
and facial landmarks reshaping (d) are extremely efficient because (b) mainly  
505 involves the inverse of a  $3 \times 3$  matrix, (c) is accelerated using a KD-tree structure,  
and (e) is simply a linear projection. While the step of facial image warping (e)  
is the most time-consuming, it is still reasonably fast given the high resolution of  
the input image. Note that the steps of facial landmarks detection (a) and facial  
image warping (e) are not the focus of this work, and they could be replaced by  
510 other faster methods in the future.

## 5. Conclusion and Future Work

We have proposed a purely unsupervised method for facial shape beautifi-  
cation, without requiring any human-annotated facial attractiveness scores for  
training. Our method has been shown to be capable of beautifying both frontal  
515 portraits and general facial images that contain a wide range of non-frontal poses  
and non-neutral expressions, which is beyond the ability of the current super-  
vised face beautification methods. On top of this, our method enables users to



Table 1: The run time of our method for beautifying a typical input facial image of resolution  $1280 \times 960$ . The steps (b)-(d) are our contributions in this work, for each of which we also list the big  $\mathcal{O}$  computational complexity (where  $n$  is the number of facial landmarks,  $m$  is the number of facial images in the dataset, and  $k$  is a small number of iteration steps required for convergence). As the steps (a) and (e) are not our work in this paper, we refer the readers to the original papers for their detailed computational complexity analysis (Le et al., 2012; Schaefer et al., 2006).

	Operation	Complexity	Run time
Step (a)	Facial Landmarks Detection	—	0.451723s
<b>Step (b)</b>	<b>Facial Operator Computation</b>	$\mathcal{O}(3n^2)$	<b>0.000117s</b>
<b>Step (c)</b>	<b>Facial Operator Optimization</b>	$\mathcal{O}(kn^2m \log_m)$	<b>0.190622s</b>
<b>Step (d)</b>	<b>Facial Landmarks Reshaping</b>	$\mathcal{O}(2n^3)$	<b>0.002041s</b>
Step (e)	Facial Image Warping	—	0.323160s
Total	The Whole Process	—	0.967839s

flexibly prescribe beautification weights so that certain facial parts can be preserved more while others can be enhanced more. As the speed of our method is interactive, it allows users to continuously customize the beautification weights until satisfactory results are obtained.

To achieve the above, we have proposed to formulate face beautification as the process of iteratively pulling the operator representation of an input face towards a local nearby density mode in the operator space of faces. Benefiting from the orthogonal projection nature of our operator representation, it frees the beautification process from the nuisance affine transformations of facial landmarks and naturally groups facial landmarks into individual facial parts for user control. It also significantly simplifies the reconstruction of beautified facial landmarks as linear projection, which is very efficient and guaranteed to be globally optimal. Due to the locality of our iterative face averaging method, the beautified version of an input face faithfully preserves the original pose and expression, which have been confirmed by many examples shown in the paper and the supplementary materials.

The main limitation of our method is that it mainly applies small modifica-

535 tions to the shape of a given face, as shown in Fig. 8. This is because modifying the facial shape to a much larger extent may introduce noticeable flaws, especially when the face is confounded with significantly non-frontal poses and non-neutral expressions. This limitation was not demonstrated in the previous methods as they exclusively handle frontal and neutral faces, which unfortunately only account for a tiny fraction of facial photos posted every day on  
540 social media.

To address the above limitation, one important future work is to find a theoretical way to decompose our orthogonal projection operator of faces into three geometrically independent variations: shape content, pose, and expression.  
545 With this decomposition, we can then safely apply larger modifications to the shape content while maintaining the pose and expression as much as possible. The previous work on the decomposition of faces (Banz & Vetter, 1999; Tena et al., 2011; Yang et al., 2012) only discovers the statistical independence of these variations, which may not suffice to maintain the realism of each individual face in the beautification process. Motivated by the recent success of (Shen et al.,  
550 2020) on semantic face editing, we are actively exploring GANs trained on our facial operator representations to obtain more reliable identity-relevant features for facial shape beautification. Based on a good geometric decomposition of our facial operator, another extension of our face beautification method is to further  
555 consider some annotated facial beauty scores while pulling an input face towards a local density mode. As we have shown in the paper, beautifying faces along the underlying face manifold is important for preserving facial characteristic features. However, our study in Fig. 12 indicates that there are around 35% of participants who do not necessarily regard those density modes of faces as  
560 significantly more attractive. This suggests that the Euclidean distance metric we use in the facial operator space is suboptimal for face beautification. As an extension, we are planning to replace the Euclidean metric with a more effective one by jointly learning from a large collection of unannotated faces and a small fraction of annotated beauty ratings (Anand et al., 2013). The recent deep  
565 learning methods of (Ge, 2018; Cakir et al., 2019) can also be extended to learn

a more effective yet beauty-aware metric for facial shape analysis. Without further modification, our method would then be able to converge to the more beauty-aware local density modes by following the new metric in our facial operator space.

570 We have recruited 12 Chinese males and 12 Chinese females in our user study and mainly reported their aggregated judgements of our face beautification results. However, previous studies have also shown that the gender and race biases of human subjects are non-negligible in the perception of facial attractiveness (Coetzee et al., 2014). Therefore, we are planning to extend our  
575 user study for more detailed data collection and analysis of such group biases in the future. We also intend to interview the future subjects on top of asking for their overall rating of face beautification results, so that we can gain more detailed subjective insights into the results.

The computational bottlenecks of our face beautification method are the  
580 steps of facial landmarks detection and facial image warping. These are not our focus in this work and can be significantly accelerated by using other faster methods, such as the one of (Zhang et al., 2014) for landmarks detection and that of (Liao et al.) for image warping.

Currently, to ease user control in face beautification, we have allowed users  
585 to customize beautification weights on the level of facial parts (e.g. the jawline and the eyes). An interesting direction of research is automatically learning these weights from high-level personal traits such as gender, age, and race.

We have formulated facial shape beautification as an iterative locality-sensitive averaging process in the face space. We have specified the locality of face averaging using a scale parameter  $\gamma$ , which has the advantage of producing minimum  
590 changes to input faces so that major facial features are well preserved. In the future, the choice of  $\gamma$  could be optimized based on user preferences towards facial attractiveness. Non-local faces could also be incorporated in the averaging process to suggest more diverse, user-satisfied face changes for beautification.

595 We have focused on the beautification of 2D facial shapes in this work. Still, our facial reshaping operator and locality-sensitive averaging method are inde-

pendent of the specific coordinate dimension, which can be seamlessly adapted for the beautification of 3D facial meshes (Liao et al., 2012).

As users prefer minimum changes to input images, we have proposed to  
600 fix the leading PCA coefficients of an input facial operator while optimizing  
the remaining. This method works well in our experiments and can be further  
enhanced using the more advanced facial attribute descriptions (Wang et al.,  
2016).

### Acknowledgements

605 This work was supported in part by the Erasmus Mundus Action 2 Pro-  
gramme (Ref: 2014-0861/001-001), the Engineering and Physical Sciences Re-  
search Council (EPSRC) (Ref: EP/M002632/1), the Royal Society  
(Ref: IES\R2\181024), the National Key Research and Development Program  
of China (Ref: 2017YFB1002702), and the National Nature Science Foundation  
610 of China (Ref: 61572058).

### References

- Anand, S., Mittal, S., Tuzel, O., & Meer, P. (2013). Semi-supervised kernel  
mean shift clustering. *IEEE transactions on pattern analysis and machine  
intelligence*, 36, 1201–1215.
- 615 Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3d faces.  
In *Proceedings of the ACM SIGGRAPH* (pp. 187–194).
- Brand, M., & Pletscher, P. (2008). A conditional random field for automatic  
photo editing. In *Proceedings of the IEEE Conference on Computer Vision  
and Pattern Recognition* (pp. 1–7).
- 620 Cakir, F., He, K., Xia, X., Kulis, B., & Sclaroff, S. (2019). Deep metric learning  
to rank. In *Proceedings of the IEEE Conference on Computer Vision and  
Pattern Recognition* (pp. 1861–1870).

- Chen, F., Xu, Y., & Zhang, D. (2014). A new hypothesis on facial beauty perception. *ACM Transactions on Applied Perception*, 11, 8:1–8:20.
- 625 Chen, F., Xu, Y., Zhang, D., & Chen, K. (2015). 2d facial landmark model design by combining key points and inserted points. *Expert systems with applications*, 42, 7858–7868.
- Ciuc, M., Capata, A., Mocanu, V., Pososin, A., Florea, C., & Corcoran, P. (2013). Automatic face and skin beautification using face detection. US  
630 Patent 8,345,114.
- Coetsee, V., Greeff, J. M., Stephen, I. D., & Perrett, D. I. (2014). Cross-cultural agreement in facial attractiveness preferences: The role of ethnicity and gender. *PloS one*, 9, e99629.
- Comaniciu, D., & Meer, P. (2002). Mean shift: a robust approach toward  
635 feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 603–619.
- Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., & Wu, C.-H. (1995). "their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical  
640 attractiveness. *Journal of Personality and Social Psychology*, 68, 261.
- Dang, L. M., Hassan, S. I., Im, S., & Moon, H. (2019). Face image manipulation detection based on a convolutional neural network. *Expert Systems with Applications*, 129, 156–168.
- Diamant, N., Zadok, D., Baskin, C., Schwartz, E., & Bronstein, A. M. (2019).  
645 Beholder-gan: Generation and beautification of facial images with conditioning on their beauty level. In *Proceedings of the IEEE International Conference on Image Processing* (pp. 739–743).
- Eisenthal, Y., Dror, G., & Ruppin, E. (2006). Facial attractiveness: Beauty and the machine. *Neural Computation*, 18, 119–142.

- 650 Gan, J., Li, L., Zhai, Y., & Liu, Y. (2014). Deep self-taught learning for facial beauty prediction. *Neurocomputing*, 144, 295–303.
- Ge, W. (2018). Deep metric learning with hierarchical triplet loss. In *Proceedings of the European Conference on Computer Vision* (pp. 269–285).
- Georgescu, B., Shimshoni, I., & Meer, P. (2003). Mean shift based clustering in  
655 high dimensions: a texture classification example. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 456–463).
- Grammer, K., & Thornhill, R. (1994). Human facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology*, 108, 233.
- 660 Guo, D., & Sim, T. (2009). Digital face makeup by example. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 73–79).
- Hassner, T. (2013). Viewing real-world faces in 3d. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3607–3614).
- 665 Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we instagram: A first analysis of instagram photo content and user types. In *Proceedings of the International AAAI conference on weblogs and social media*.
- Joshi, N., Matusik, W., Adelson, E. H., & Kriegman, D. J. (2010). Personal photo enhancement using example images. *ACM Transactions on Graphics*,  
670 29, 12:1–12:15.
- Kemelmacher-Shlizerman, I., Shechtman, E., Garg, R., & Seitz, S. M. (2011). Exploring photobios. *ACM Transactions on Graphics*, 30, 61:1–61:10.
- Laurentini, A., & Bottino, A. (2014). Computer analysis of face beauty: A survey. *Computer Vision and Image Understanding*, 125, 184–199.

- 675 Le, V., Brandt, J., Lin, Z., Bourdev, L., & Huang, T. S. (2012). Interactive facial  
feature localization. In *Proceedings of the European Conference on Computer  
Vision* (pp. 679–692).
- Leyvand, T., Cohen-Or, D., Dror, G., & Lischinski, D. (2008). Data-driven  
enhancement of facial attractiveness. *ACM Transactions on Graphics*, 27,  
680 38:1–38:9.
- Li, J., Xiong, C., Liu, L., Shu, X., & Yan, S. (2015). Deep face beautification.  
In *Proceedings of the ACM international conference on Multimedia* (pp. 793–  
794).
- Li, Z., & Tang, Y. (2018). Comparative density peaks clustering. *Expert Systems  
with Applications*, 95, 236–247.  
685
- Liao, J., Lima, R. S., Nehab, D., Hoppe, H., Sander, P. V., & Yu, J. (). Au-  
tomating image morphing using structural similarity on a halfway domain.  
*ACM Transactions on Graphics*, 33, 168:1–168:12.
- Liao, Q., Jin, X., & Zeng, W. (2012). Enhancing the symmetry and propor-  
tion of 3d face geometry. *IEEE Transactions on Visualization and Computer  
Graphics*, 18, 1704–1716.  
690
- Maleš, L., Marčetić, D., & Ribarić, S. (2019). A multi-agent dynamic system for  
robust multi-face tracking. *Expert Systems with Applications*, 126, 246–264.
- Pasupa, K., Sunhem, W., & Loo, C. K. (2019). A hybrid approach to building  
face shape classifier for hairstyle recommender system. *Expert Systems with  
Applications*, 120, 14–32.  
695
- Schaefer, S., McPhail, T., & Warren, J. (2006). Image deformation using moving  
least squares. *ACM Transactions on Graphics*, 25, 533–540.
- Scherbaum, K., Ritschel, T., Hullin, M., Thormählen, T., Blanz, V., & Seidel,  
700 H.-P. (2011). Computer-suggested facial makeup. . 30, 485–492.

- Schmid, K., Marx, D., & Samal, A. (2008). Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratios. *Pattern Recognition*, 41, 2710–2717.
- Seidman, G., & Miller, O. S. (2013). Effects of gender and physical attractiveness on visual attention to facebook profiles. *Cyberpsychology, Behavior, and Social Networking*, 16, 20–24.
- Shen, Y., Gu, J., Tang, X., & Zhou, B. (2020). Interpreting the latent space of gans for semantic face editing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 9243–9252).
- Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., & Samuels, C. (1998). Newborn infants prefer attractive faces. *Infant Behavior and Development*, 21, 345–354.
- Tena, J. R., De la Torre, F., & Matthews, I. (2011). Interactive region-based linear 3d face models. *ACM Transactions on Graphics*, 30, 76:1–76:10.
- Wang, J., Cheng, Y., & Schmidt Feris, R. (2016). Walk and learn: Facial attribute representation learning from egocentric video and contextual data. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2295–2304).
- Winston, J. S., O’Doherty, J., Kilner, J. M., Perrett, D. I., & Dolan, R. J. (2007). Brain systems for assessing facial attractiveness. *Neuropsychologia*, 45, 195–206.
- Yang, F., Bourdev, L., Shechtman, E., Wang, J., & Metaxas, D. (2012). Facial expression editing in video using a temporally-smooth factorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 861–868).
- Yang, F., Wang, J., Shechtman, E., Bourdev, L., & Metaxas, D. (2011). Expression flow for 3d-aware face component transfer. *ACM Transactions on Graphics*, 30, 60:1–60:10.



- 730 Zhang, D., Zhao, Q., & Chen, F. (2011). Quantitative analysis of human facial beauty using geometric features. *Pattern Recognition*, *44*, 940–950.
- Zhang, J., Shan, S., Kan, M., & Chen, X. (2014). Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment. In *Proceedings of the European Conference on Computer Vision* (pp. 1–16).
- 735 Zhang, L., Shum, H. P. H., Liu, L., Guo, G., & Shao, L. (2019). Multiview discriminative marginal metric learning for makeup face verification. *Neurocomputing*, *333*, 339–350.
- Zhang, L., Zhang, D., Sun, M.-M., & Chen, F.-M. (2017). Facial beauty analysis based on geometric feature: Toward attractiveness assessment application. *Expert Systems with Applications*, *82*, 252–265.